Improving Medical Surveillance through Fusing Disparate Evidence

Jeffrey Lin[1], Howard Burkom[1], Andrew B. Feldman[1], Sean Murphy[1], Yevgeniy Elbert[2], Shilpa Hakre[2], Steven Babin[1]

[1]The Johns Hopkins University Applied Physics Laboratory

[2] Walter Reed Army Institute for Research

Scientific Conference on Chemical and Biological Defense Research

Nov. 15, 2004

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
15 NOV 2004	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Improving Medical Surveillance through Fusing Disparate Evidence	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **The Johns Hopkins University Applied Physics Laboratory**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001849, 2004 Scientific Conference on Chemical and Biological Defense Research. Held in Hunt Valley, Maryland on 15-17 November 2004 . , The original document contains color images.

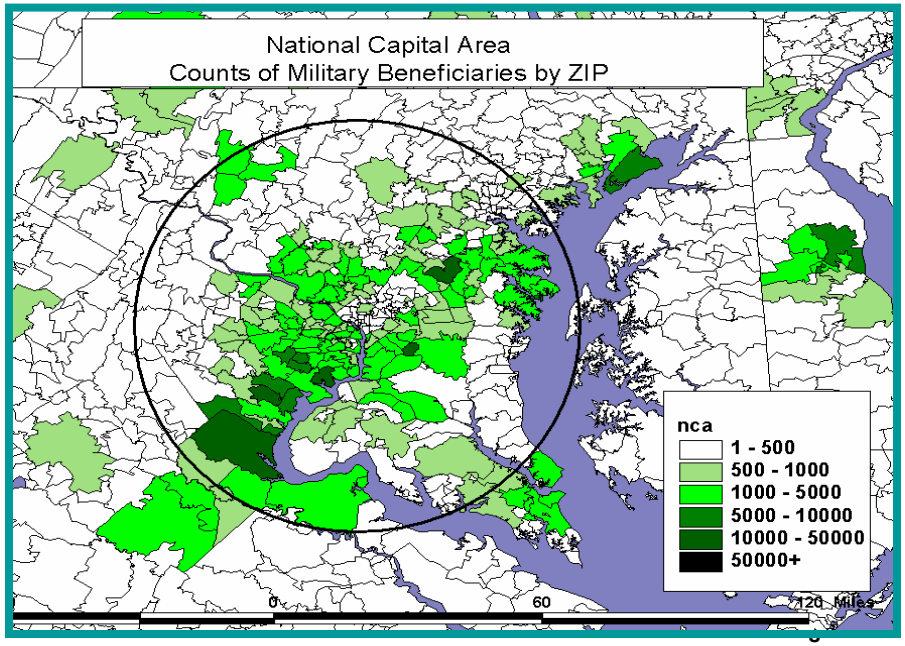
14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**24**	

ESSENCE Biosurveillance Systems

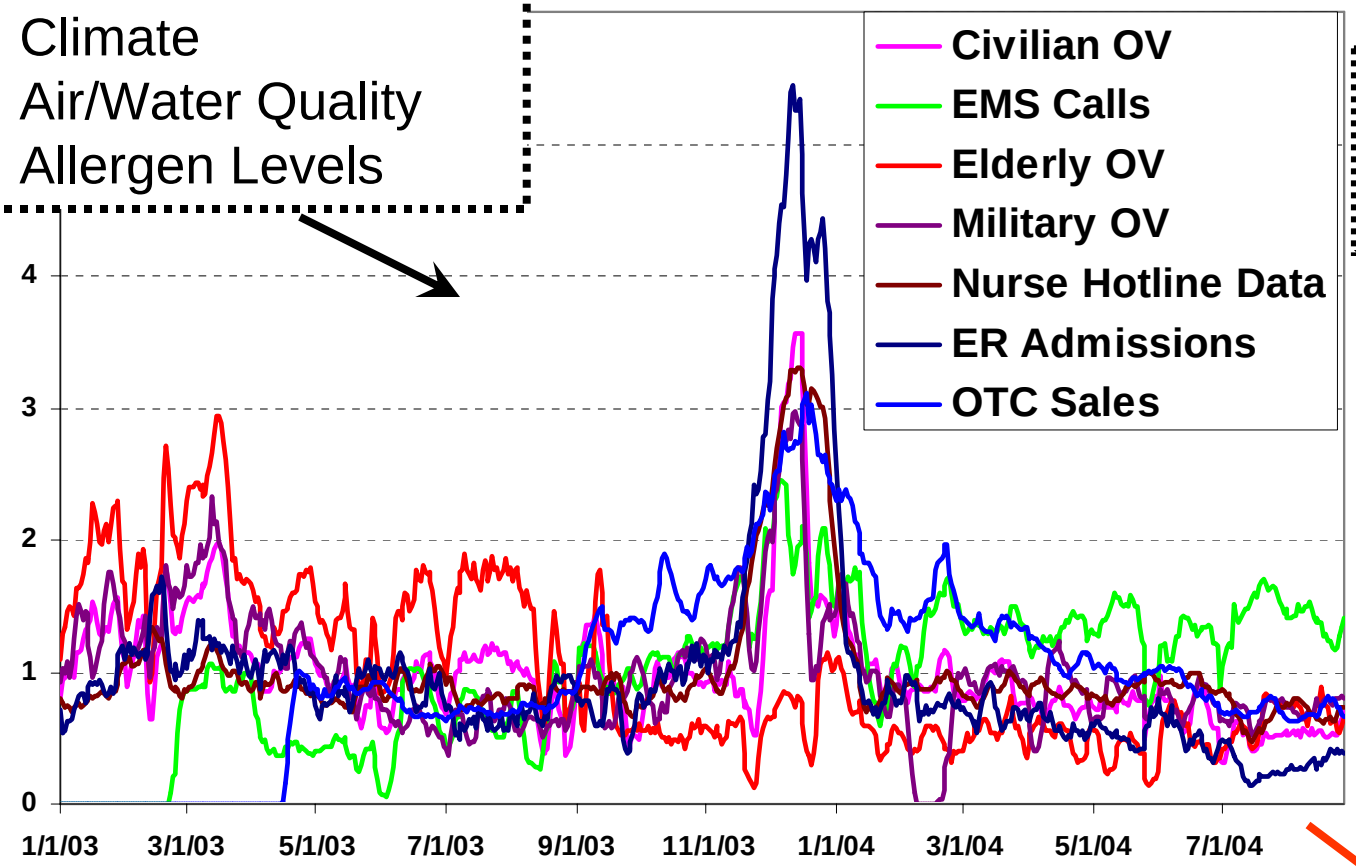
- **ESSENCE**: An **E**lectronic **S**urveillance **S**ystem for the **E**arly **N**otification of **C**ommunity-based **E**pidemics
- Monitoring health care data
  - ~800 military treatment facilities since Sept. 2001
  - 12 major metropolitan civilian areas
- Evaluating data sources
  - Civilian physician visits
  - OTC pharmacy sales
  - Prescription sales
  - Nurse hotline/EMS data
  - Absentee rate data
- Developing & implementing alerting algorithms



Envisioned: Decisions Based on Disparate Evidence

Environmental Data:
- Climate
- Air/Water Quality
- Allergen Levels

Syndromic Time Series



Biosensor Data

Integrated Threat Assessment

Statistical Tools

- Aberration detection algorithms
  - Data modeling: multivariate regression
    - Covariates: Holiday, post-holiday, trend, provider count,…
  - Statistical process control

    EWMA, CUSUM charts
- Combining data sources
  - Multiple univariate: combine p-values
  - Multivariate: Hotelling's T^2 variants: MEWMA, …

Elements of Data Fusion Problem

- Evidence disparate in scale, variability, specificity, timeliness
  - *syndromic*: ED data specific, possibly late; OTC data nonspecific, potentially timely
  - *sensor*: sparse spatial coverage; data gaps
- Informatics issues
  - Differential lags in signal effect, reporting
  - Data dropouts
- Differential background characterization
- Differential signal characterization
- Differential information value (relevance)

Bayes Belief Net (BBN) Umbrella

- Graphical representation of conditional dependencies
- Inclusion of disparate evidence types
  - Continuous/discrete data or derived probabilities
  - Expert/heuristic knowledge
- Can weight statistical hypothesis test evidence using heuristics – not restricted to fixed p-value thresholds
- Can exploit advances in data modeling, multivariate anomaly detection
- Modularity in data fusion approach
- Management of missing data
- Can model
  - Personal weighting of evidence
  - Lags in data availability or reporting

Model Building – Clinical Models Exist

Inhalational anthrax … a biphasic clinical illness …
1-to 4-day initial phase of malaise, fatigue, fever, myalgias, and nonproductive cough, followed by a fulminant [sudden and severe] phase of respiratory distress, cyanosis, and diaphoresis [sweating]. Death follows the onset of the fulminant phase in 1 to 2 days.

John A. Jernigan, et al., "Bioterrorism-Related Inhalational Anthrax: The First 10 Cases Reported in the United States," *Emerging Infectious Diseases*, Vol. 7, No. 6, November-December 2001

Data from the Sverdlovsk outbreak indicate a modal incubation time of approximately 10 days for inhalational anthrax. However, the onset of symptoms occurred up to six weeks after the reported date of exposure. Such long incubation times presumably reflect the ability of viable anthrax spores to remain in the lungs for many days. Longer incubation periods may be associated with smaller inocula.

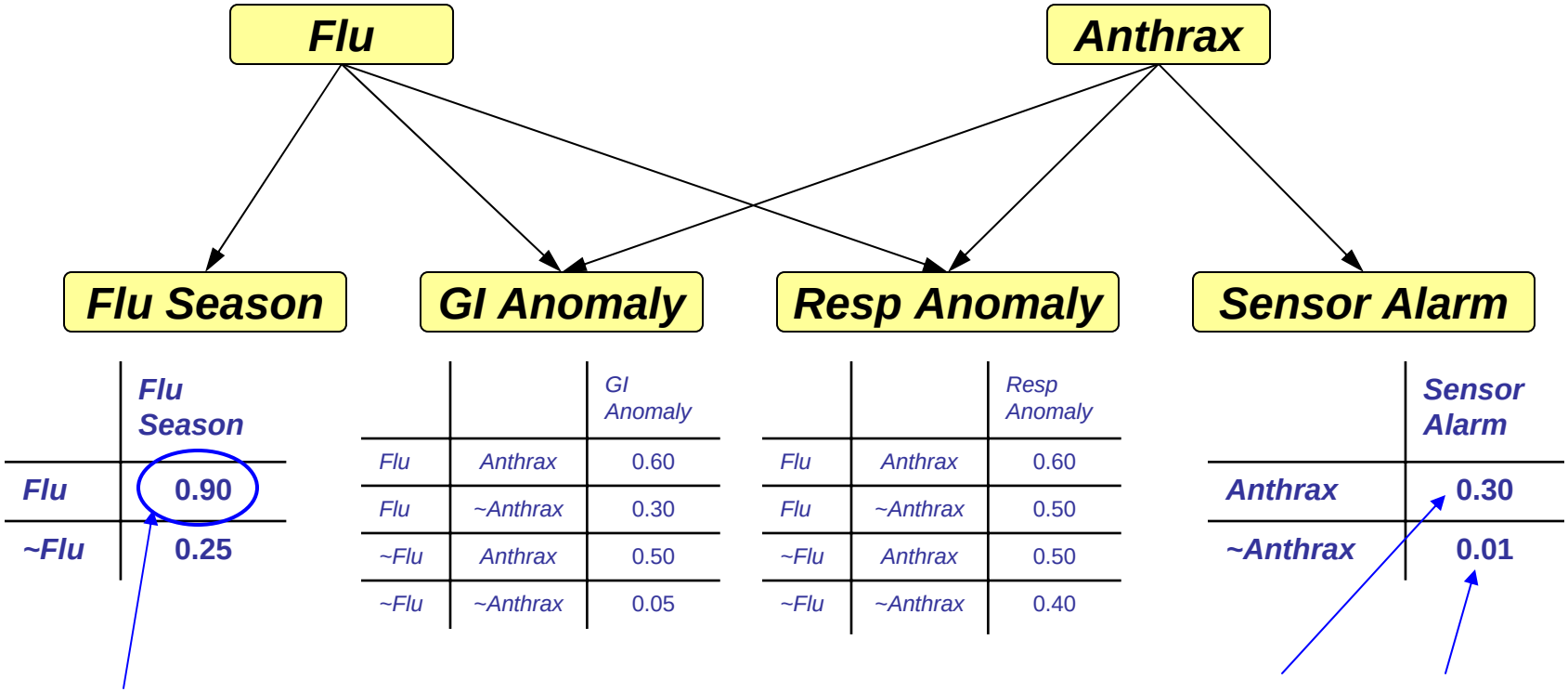
Terry C. Dixon, B.S., et al., "Anthrax," *NEJM* , Volume 341:815-826, Number 11, September 9, 1999

Example Bayes Network (1)

Prior Probabilities

P(*Flu Outbreak Occurring*) = 0.05

P(*Anthrax Outbreak Occurring*) = 0.001



	Flu Season
Flu	0.90
~Flu	0.25

		GI Anomaly
Flu	Anthrax	0.60
Flu	~Anthrax	0.30
~Flu	Anthrax	0.50
~Flu	~Anthrax	0.05

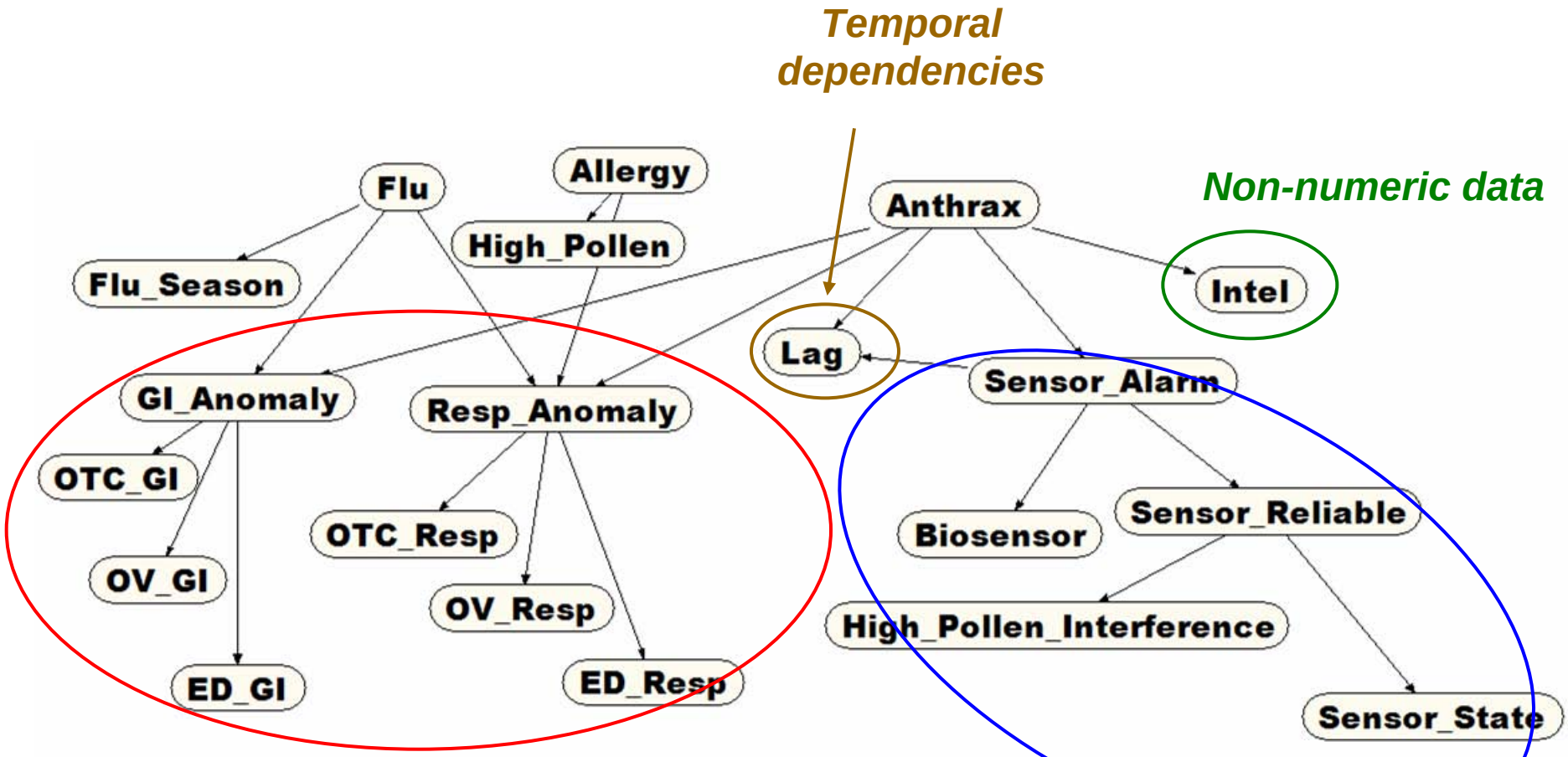
		Resp Anomaly
Flu	Anthrax	0.60
Flu	~Anthrax	0.50
~Flu	Anthrax	0.50
~Flu	~Anthrax	0.40

	Sensor Alarm
Anthrax	0.30
~Anthrax	0.01

P(*Flu Season* | *Flu Outbreak Occurring*) = 0.90

Effective Sensor PD and PFA

Notional Bayes Network for Event Classification



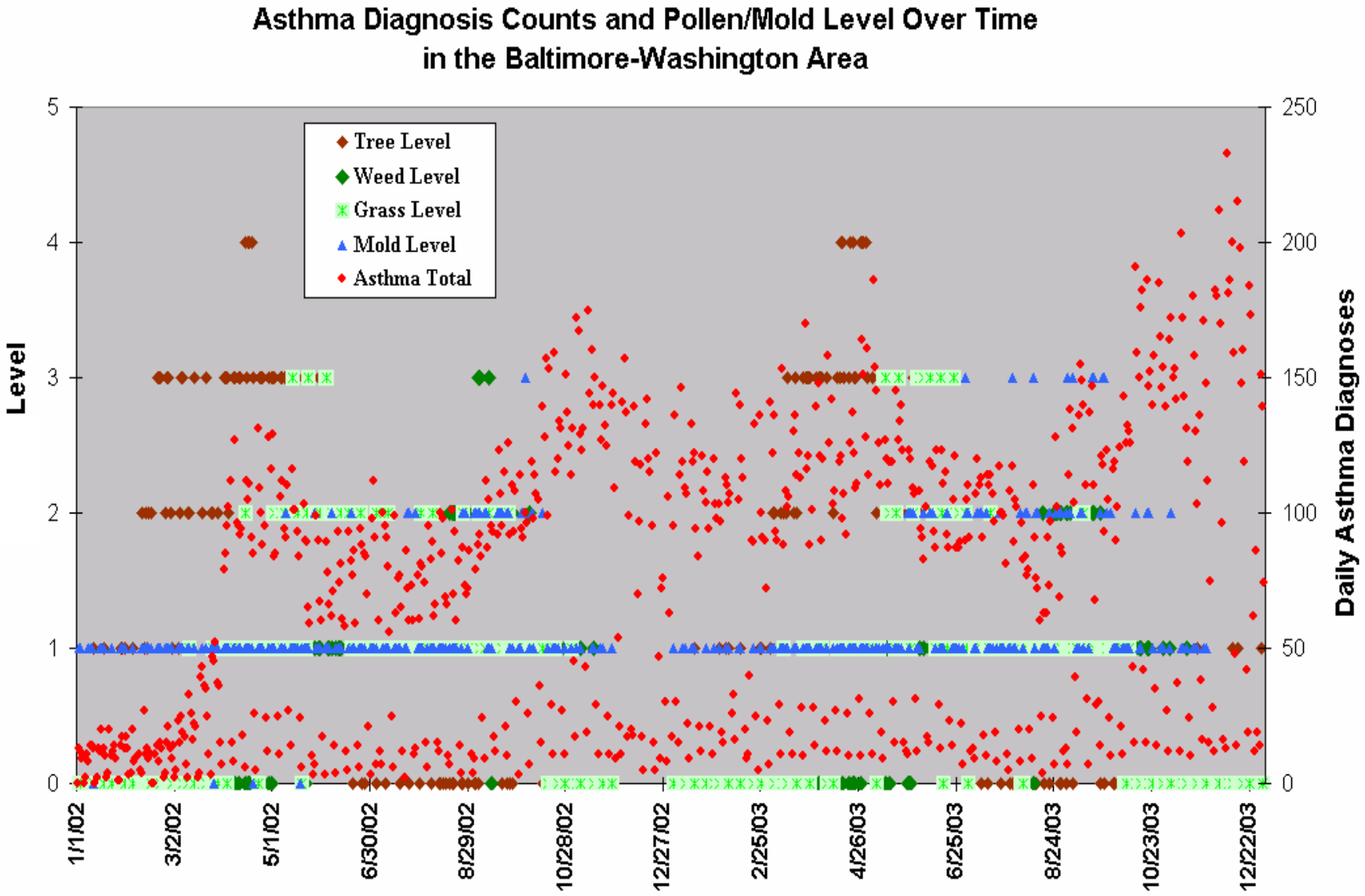
Application to Asthma Flare-ups

- Availability of practical, verifiable data:
  - For “truth data”: daily clinical diagnosis counts
  - For “evidence”: daily environmental, syndromic data
- Known asthma triggers with complex interaction
  - Air quality (EPA data)
    - Concentration of particulate matter, allergens
    - Ozone levels
  - Temperature (NOAA data)
  - Viral infections (Syndromic data)
- Evidence from combination of expert knowledge, historical data

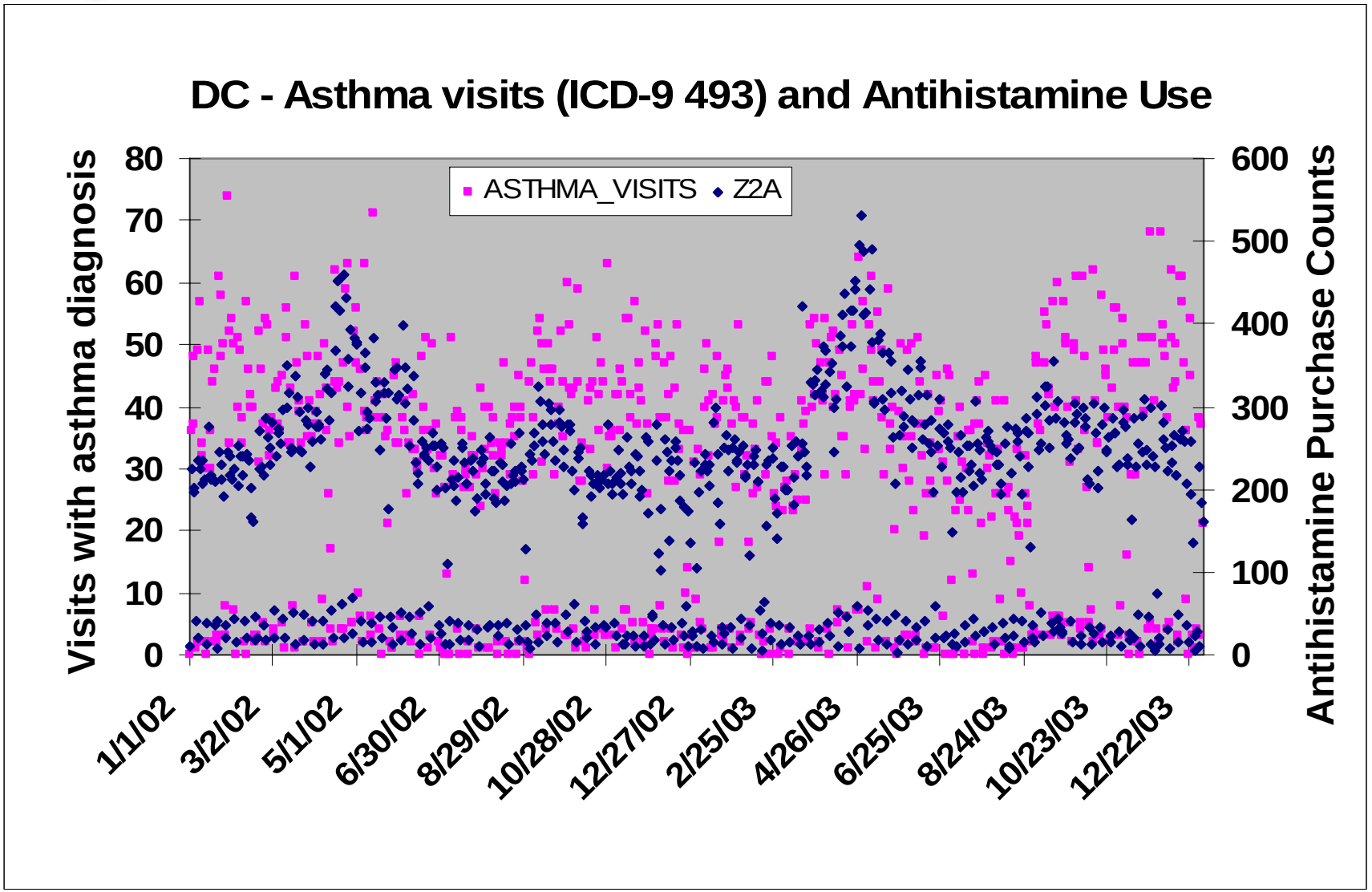
Asthma Triggers: Expert Evidence

- **Ozone:**
  - Burnett et al, 1994;
  - Sartor et al, 1995;
  - Stern et al, 1994;
  - Stieb et al, 1996;
  - Zhang et al, 2004 and others.
- **Particulate Matter (PM):**
  - Anderson et al, 2001;
  - Chuersuwan et al 2000;
  - Leaderer et al, 2003;
  - Howel et al, 2001;
  - Norris et al, 1999;
  - Ward and Ayres, 2004 and others.
- **Allergens:**
  - Solomon 2002;
  - Taylor et al 2002;
  - Ziska et al, 2003 and others,
- **Viral Infections:**
  - Hegele, 1999;
  - Cohen and Castro, 2003;
  - Lemanske, 2003 and others;
- **Cold Weather:**
  - Anderson et al, 2001;
  - Jamason et al, 1997;
  - Packe and Ayres, 1985;
  - Sartor et al, 1995;
  - Schachter et al, 1981, others.

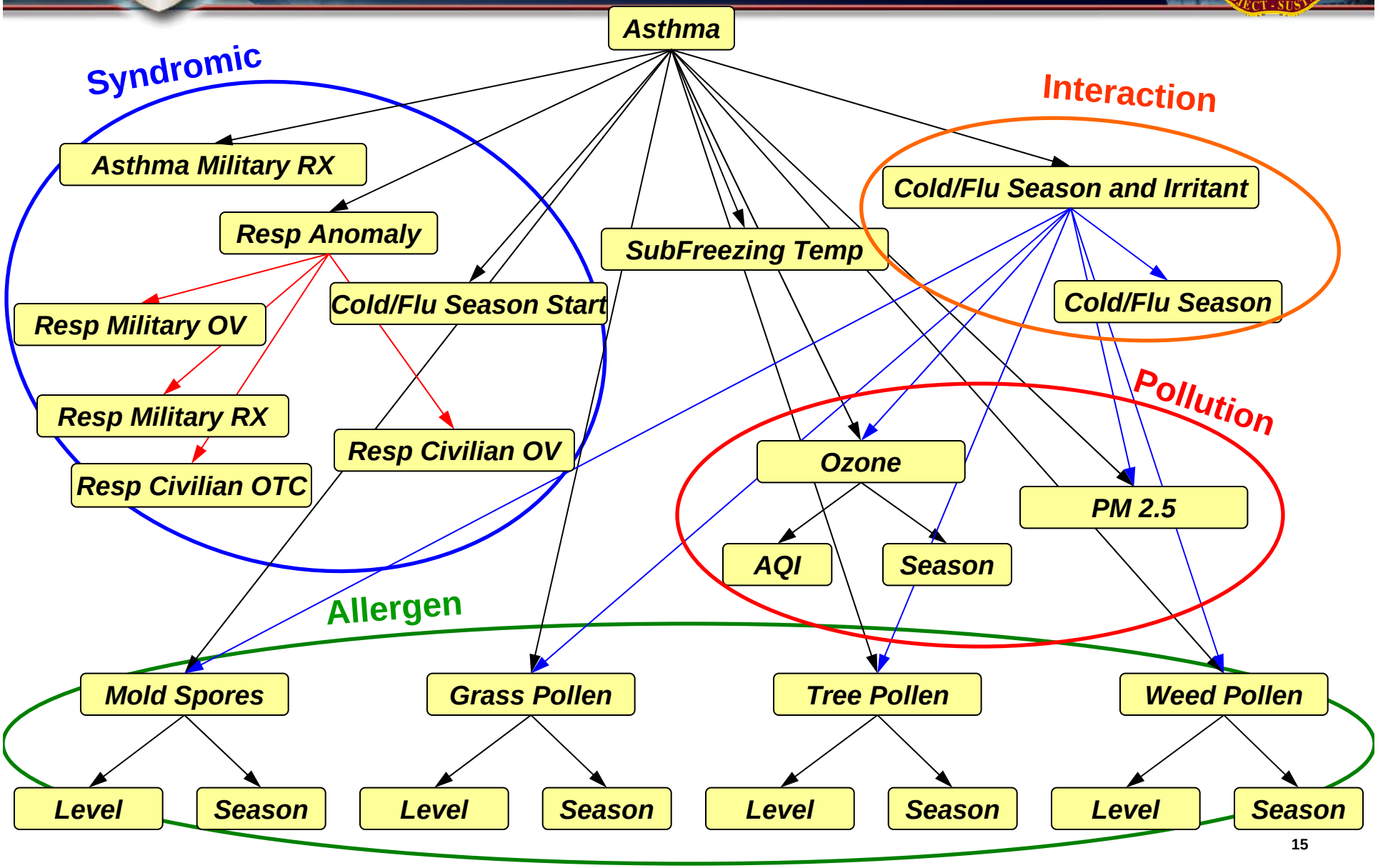
Environmental Evidence: Allergen Levels and Diagnosis Counts



Syndromic Evidence: OTC Sales and Diagnosis Counts



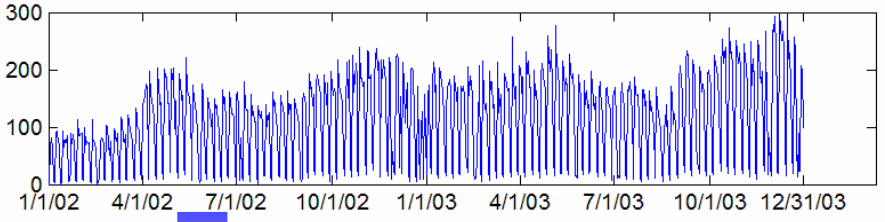
Structure of BBN Model for Asthma Flare Ups



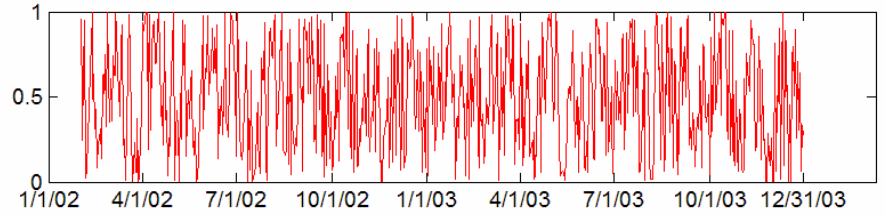
Data Flow Diagram

1. All NCR county military and civilian asthma and provider counts are totaled.
2. Regression algorithm seeks ‘anomalies’ taking into account:
   - Day of week
   - Holidays
   - Data trends
3. Regression output is rescaled using a sigmoidal function designed to “stretch” out the high end of the regression output.
4. Output > 0.9 are chosen as flare up ‘seeds’ and extended three days before and 1 day after to generate “truth.”

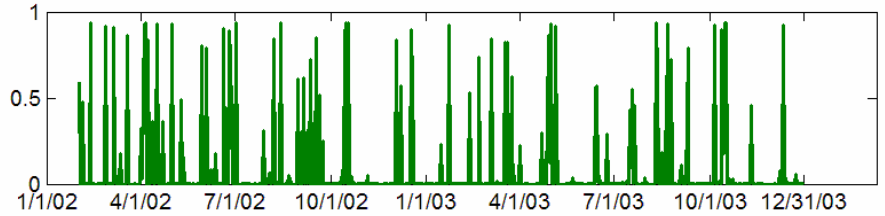
1. Total NCR Asthma and Provider Count



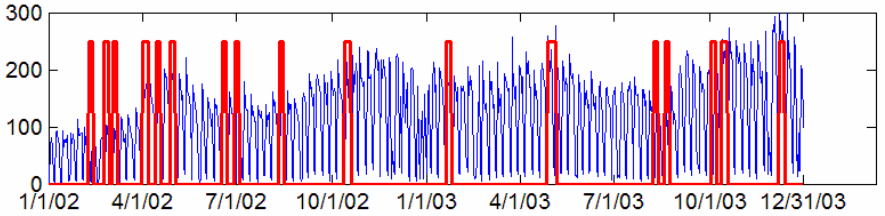
2. Regression



3. Probability Map



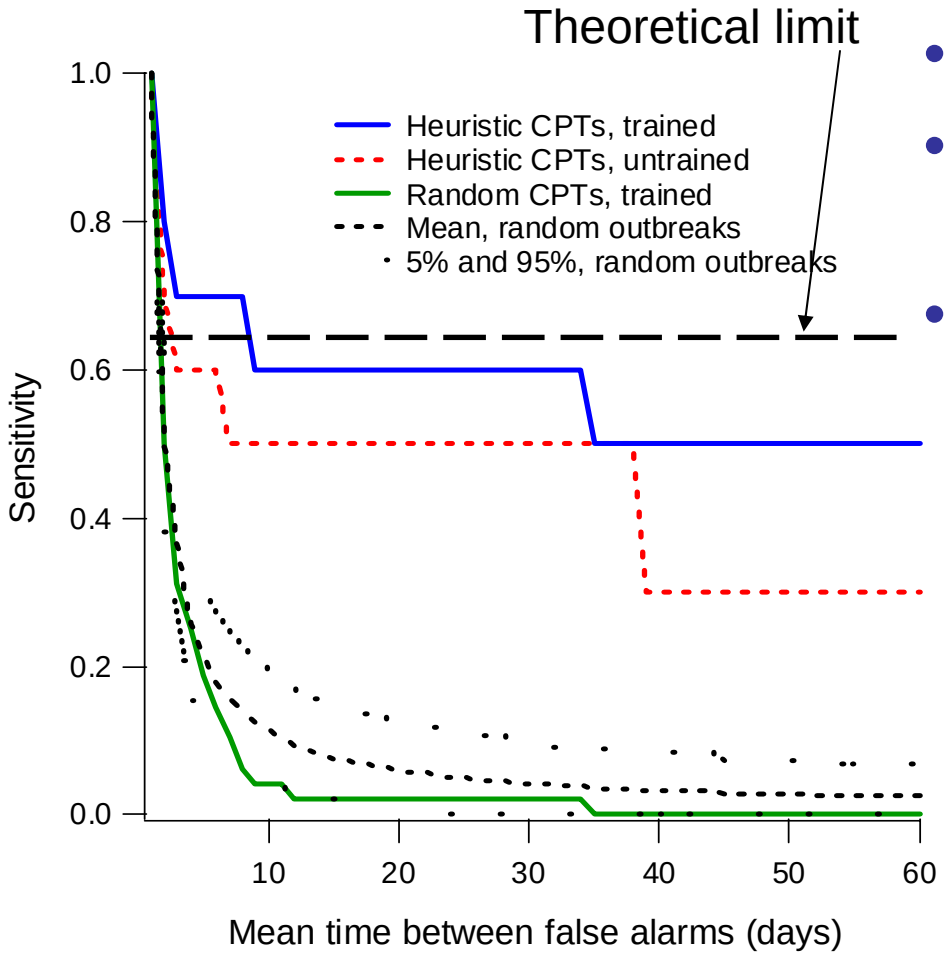
4. Unbiased Asthma Flare Ups



Bayesian Network Learning

- Structure Learning
  - Determining nodes, edges of graph: what are the effective relationships (cond. dependencies) among data types, other nodes? (not automated: only heuristic structure used)
- Parameter Learning
  - Maximum Likelihood Estimation (MLE): compute CPTs that best explain data in a "brute force" frequency density sense
    - Then $\text{Prob}_{\mathbf{MLE}}(\text{data}) = \text{Prob}(\text{data} \mid \text{MLE CPTs})$
  - Maximum *A Posteriori* (MAP): compute CPTs that best explain data *given prior CPT estimates*, along with weights
    - Then $\text{Prob}_{\mathbf{MAP}}(\text{data}) = \text{Prob}(\text{data} \mid \text{MAP CPTs})$

Asthma Detector Results

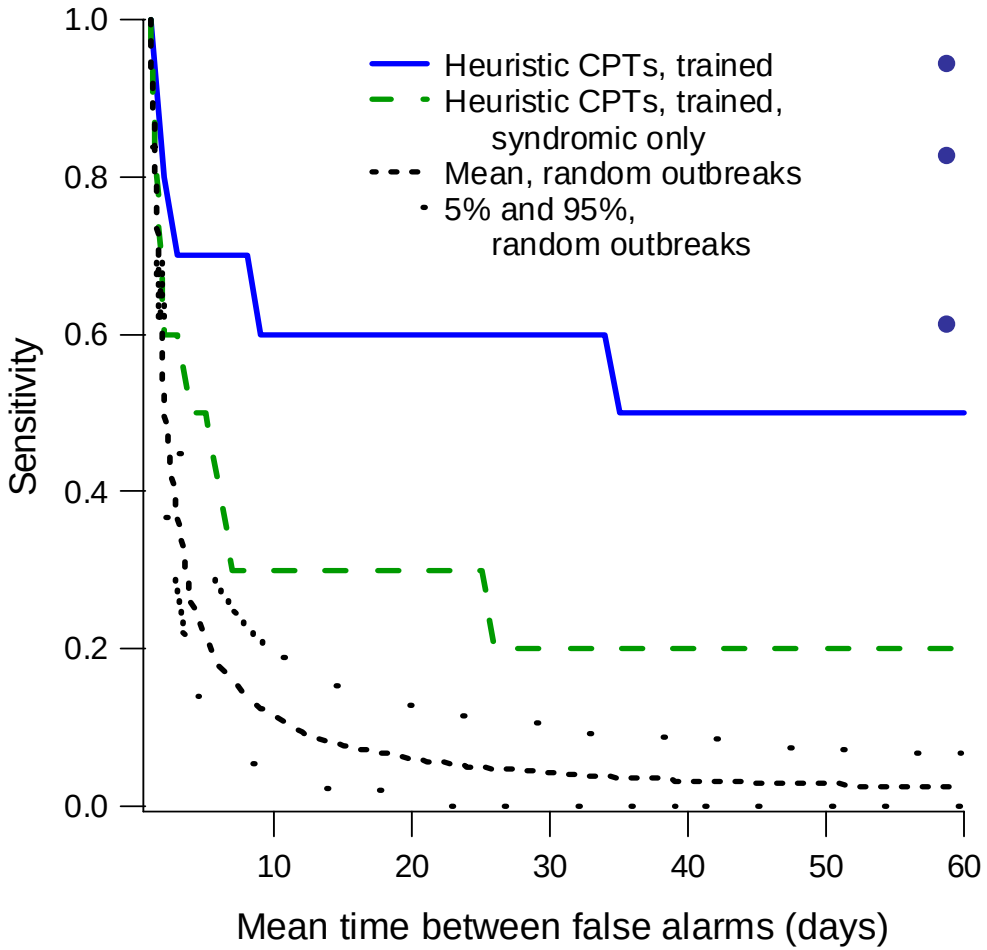


- ROC curve for 2002
- All NCR, military and civilian
- Asthma "outbreaks"
  - 10 (auto) identified
  - 5 day windows

All-heuristic BBN performs very well

All bio-terror networks require heuristic parameters

Asthma Detector Results



- ROC curve for 2002
- All NCR, military and civilian
- Asthma "outbreaks"
  - 10 (auto) identified
  - 5 day windows

Fusion of sensor data critical to sensitivity

Scalability

- Inferencing/learning with BBNs is NP-hard
- Heuristics severely constrain problem
  - Data is aggregated to increase SNR
  - Only select data is used as evidence
  - Modularity of structure allows approximations that reduce computations
- Mean-field approximations

Conclusions

- As a classifier, untrained heuristic-only BBN significantly outperformed
  - BBN against same flare-ups with randomized days of occurrence
  - BBN trained with data by MLE from random initial CPTs
- MLE training improved heuristic-only BBN performance across range of practical false alarm rates
- Sensitivity analysis using ROC curve analysis can reveal contributions of individual data sources; fusion with sensor data outperformed syndromic alone
- BBN modeling “works”, but for effective real-world performance, development of tools for improving graph structure, parameter learning, and prior probabilities is needed along with underlying data analysis

Ongoing efforts

- Application-related
  - Obtain & analyze biosensor data for background characterization
  - Develop cond. prob. tables for inclusion in BBN
- BBN Learning-related
  - Evaluate & compare parameter learning approaches
  - Test model variations
- Validation-related (with improved datasets)
  - Temporal cross-validation: *e.g.* application of 2003-based CPTs to 2004
  - Spatial cross-validation: *e.g.* application of NCR-data-based CPTS to San Diego, other areas

BACKUPS

Bayes' Rule in Surveillance Context

$$\text{Posterior Probability} = \frac{\text{Conditional Likelihood} * \text{Prior Probability}}{\text{Marginal Likelihood}}$$

Example:

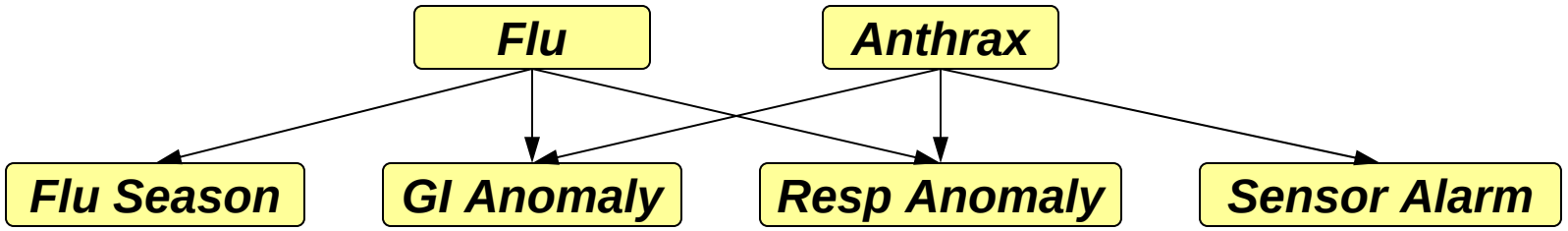
Posterior probability = Prob (anthrax attack | biosensor alert)

Conditional likelihood = Prob (biosensor alert | anthrax attack)

Prior probability = Prob (anthrax attack)

Marginal likelihood = Prob (biosensor alert)

Example Bayes Network (2)



Evidence

Flu Season	GI Anomaly	Resp Anomaly	Sensor Alarm
Flu Season	GI Anomaly	Resp Anomaly	Sensor Alarm
Flu Season	GI Anomaly	Resp Anomaly	Sensor Alarm
Flu Season	GI Anomaly	Resp Anomaly	Sensor Alarm

Posterior probabilities

P(*Flu* | *Evidence*)		P(*Anthrax* | *Evidence*)
0.70	>>	0.0023
0.67	>>	0.09
0.08	>	0.005
0.07	<	0.17